

Comparing cosmic shear measures

Optimizing the Information content of cosmic shear data vectors

Tim Eifler¹, Martin Kilbinger^{1,2}, Peter Schneider¹

1)Argelander-Institut für Astronomie, Universität Bonn, Auf dem Hügel 71, D-53121 Bonn, Germany

2) Institut d’Astrophysique de Paris, 98bis boulevard Arago, F-75014 Paris, France

ABSTRACT

Aims. We introduce an optimized data vector of cosmic shear measures (\mathcal{N}). This data vector has high information content, is not sensitive against B-mode contamination and only shows small correlation between data points of different angular scales.

Methods. We show that a data vector of the two-point correlation function (2PCF), hereafter denoted as ξ , in general contains more information on cosmological parameters compared to a data vector of the aperture mass dispersion, hereafter referred to as $\langle M_{\text{ap}}^2 \rangle$. Reason for this is the fact that $\langle M_{\text{ap}}^2 \rangle$ lacks the information of the convergence power spectrum (\mathcal{P}_κ) on large angular scales, which is contained in ξ . Therefore we create a combined data vector \mathcal{N} , which retains the advantages of $\langle M_{\text{ap}}^2 \rangle$ and in addition is also sensitive to the large-scale information of \mathcal{P}_κ . We compare the information content of the three data vectors by performing a detailed likelihood analysis and use ray-tracing simulations to derive the covariance matrices. In the last part of the paper we contaminate all data vectors with B-modes on small angular scales and examine their robustness against this contamination.

Results. The combined data vector \mathcal{N} strongly improves constraints on cosmological parameters compared to $\langle M_{\text{ap}}^2 \rangle$. Although, in case of a pure E-mode signal the information content of ξ is higher, in the more realistic case where B-modes are present the 2PCF data vector is strongly contaminated and yields biased cosmological parameter estimates. \mathcal{N} shows to be robust against this contamination. Furthermore the individual data points of \mathcal{N} show a much smaller correlation compared to ξ , leading to an almost diagonal covariance matrix.

Key words. cosmology: theory - gravitational lensing - large-scale structure of the Universe - methods: statistical

1. Introduction

Weak gravitational lensing by the Large-Scale Structure (LSS), called cosmic shear, has become a valuable for cosmology. Since the first detection of cosmic shear in 2000 (Bacon et al. 2000; Kaiser et al. 2000; van Waerbeke et al. 2000; Wittman et al. 2000), several surveys have been carried out with various depth and width. The latest results show the ability of cosmic shear to constrain cosmological parameters, in particular σ_8 (e.g. van Waerbeke et al. 2005; Semboloni et al. 2006; Hoekstra et al. 2006; Schrabback et al. 2007; Hatterscheidt et al. 2007; Massey et al. 2007). These constraints will improve even more in the near future, when the VST Kilo-degree survey will cover an area of 1700 deg^2 with a depth of 15 galaxies per arc minute², enabling us to estimate the shear signal with less than 1 % statistical error. This improvement of measuring cosmic shear should go along with an optimization of the data analysis. It is desirable to extract as much information as possible from the observational data and to derive constraints free of any contamination. Currently, most cosmic shear surveys only

consider second-order shear statistics, for which all information is contained in the power spectrum of the convergence (\mathcal{P}_κ). Although \mathcal{P}_κ is not directly measureable, it is linearly related to second-order cosmic shear measures (e.g. the two-point correlation function and the aperture mass dispersion), which can be estimated from the distorted ellipticities of the observed galaxies. More precisely, all second-order measures are filtered versions of \mathcal{P}_κ and the corresponding filter functions determine how the information content of \mathcal{P}_κ is sampled. It is the intention of this paper to compare several data vectors of cosmic shear measures and to create an optimal data vector with high information content, largely uncorrelated data points and only little sensitivity to a possible B-mode contamination. We first compare the information content of the two-point correlation function (2PCF) and aperture mass dispersion ($\langle M_{\text{ap}}^2 \rangle$). We prove a general statement that a data vector consisting of 2PCF data points (ξ) always gives tighter constraints on cosmological models compared to a data vector consisting of $\langle M_{\text{ap}}^2 \rangle$ data points ($\langle M_{\text{ap}}^2 \rangle$) and we confirm this by a likelihood analysis of ray-tracing simulations. This result cannot surprise since the 2PCF integrates over all scales of \mathcal{P}_κ and especially collects

information on large angular scales which is not taken into account by the aperture mass dispersion. Nevertheless $\langle M_{\text{ap}}^2 \rangle$ has important advantages. First, it can be used to separate E-modes and B-modes (Crittenden et al. 2002; Schneider et al. 2002b), more precisely $\langle M_{\text{ap}}^2 \rangle$ is sensitive to E-modes only. Second, due to its narrow filter function it provides highly localized information on \mathcal{P}_κ , implying that two different $\langle M_{\text{ap}}^2 \rangle$ data points are much less correlated compared to the 2PCF. Third, $\langle M_{\text{ap}}^2 \rangle$ can be easier extended to higher-order statistics (Schneider et al. 2005). These advantages are valuable and should be maintained, but the information content should be improved. Hence, we extend the $\langle M_{\text{ap}}^2 \rangle$ data vector by one data point of $\xi_+(\theta_0)$, which provides the large-scale information of \mathcal{P}_κ and call this new data vector \mathcal{N} . We perform a likelihood analysis for \mathcal{N} , examine its ability to constrain cosmological parameters and compare it to the two aforementioned data vectors.

This paper is organized as follows: Sect. 2 summarizes the basic theoretical background of 2PCF and $\langle M_{\text{ap}}^2 \rangle$. Next we compare the information content of these two second-order measures and introduce the improvement to the $\langle M_{\text{ap}}^2 \rangle$ data vector (Sect. 3). We perform a detailed likelihood analysis for the three data vectors and present the results in Sect. 4 and Sect. 5. In Sect. 6 we contaminate our shear data vectors with B-modes and again perform the likelihood analysis to investigate how significantly each data vector is influenced. Finally in Sect. 7 we discuss the results and give our conclusions. One final remark should be made on the notation. ξ and $\langle M_{\text{ap}}^2 \rangle$ denote theoretical quantities calculated from a given power spectrum, whereas $\hat{\xi}$ and \mathcal{M} are estimators obtained by averaging over many data points inside a bin. Vectors and matrices are written in bold font.

2. Two-point statistics of cosmic shear

In this section we briefly review the basics of two-point statistics, definitions of shear estimators and corresponding covariances, closely following the paper of Schneider et al. (2002a). For more details on these topics the reader is referred to Bartelmann & Schneider (2001) or, more recently, Schneider (2006).

2.1. Two-point correlation function and aperture mass dispersion

To measure the shear signal we define $\boldsymbol{\theta}$ as the connecting vector of two points and specify tangential and cross-component of the shear γ as

$$\gamma_t = -\text{Re}(\gamma e^{-2i\varphi}) \quad \text{and} \quad \gamma_x = -\text{Im}(\gamma e^{-2i\varphi}), \quad (1)$$

where φ is the polar angle of $\boldsymbol{\theta}$. The 2PCFs depend only on the absolute value of $\boldsymbol{\theta}$ and are defined as

$$\xi_\pm(\theta) = \langle \gamma_t \gamma_t \rangle(\theta) \pm \langle \gamma_x \gamma_x \rangle(\theta). \quad (2)$$

The observed shear field can be decomposed into a gradient component (called E-mode) and a curl component (B-mode) (Crittenden et al. 2002; Schneider et al. 2002b). B-modes are considered to be a contamination of the pure lensing signal,

due to noise or unresolved systematics. The limited validity of the Born approximation (Jain et al. 2000) or redshift source clustering (Schneider et al. 2002b) can also create B-modes, although these effects are small. Intrinsic alignment of source galaxies is another possible explanation. Predictions about the impact of this effect differ, anyway it can be overcome when using photometric redshifts (King & Schneider 2003). For the case of a general shear field consisting of E- and B-modes, the convergence is also complex, $\kappa = \kappa_E + i\kappa_B$, and it can be related to the shear (Kaiser & Squires 1993) by

$$\kappa_E(\theta) + i\kappa_B(\theta) = \frac{1}{\pi} \int d^2\theta' \mathcal{D}^*(\theta - \theta') \gamma(\theta'), \quad (3)$$

with

$$\mathcal{D}^*(\theta) = \frac{\theta_2^2 - \theta_1^2 + 2i\theta_1\theta_2}{|\theta|^4}. \quad (4)$$

The power spectra of E-mode and B-mode can be defined (Schneider et al. 2002b) using the Fourier transform of κ

$$\langle \hat{\kappa}_E(\boldsymbol{\ell}) \hat{\kappa}_E^*(\boldsymbol{\ell}') \rangle = (2\pi)^2 \delta^{(2)}(\boldsymbol{\ell} - \boldsymbol{\ell}') \mathcal{P}_E(\ell), \quad (5)$$

$$\langle \hat{\kappa}_B(\boldsymbol{\ell}) \hat{\kappa}_B^*(\boldsymbol{\ell}') \rangle = (2\pi)^2 \delta^{(2)}(\boldsymbol{\ell} - \boldsymbol{\ell}') \mathcal{P}_B(\ell), \quad (6)$$

$$\langle \hat{\kappa}_E(\boldsymbol{\ell}) \hat{\kappa}_B^*(\boldsymbol{\ell}') \rangle = (2\pi)^2 \delta^{(2)}(\boldsymbol{\ell} - \boldsymbol{\ell}') \mathcal{P}_{EB}(\ell), \quad (7)$$

with $\delta^{(2)}(\boldsymbol{\ell})$ as the two-dimensional Dirac delta distribution. The cross power spectrum \mathcal{P}_{EB} is expected to vanish for a statistically parity-invariant shear field. Note that \mathcal{P}_E can be related to the power spectrum of density fluctuations \mathcal{P}_δ via Limber's equation (Kaiser 1992, 1998)

$$\mathcal{P}_E(\ell) = \frac{9H_0^4 \Omega_m^2}{4c^4} \int_0^{w_h} \frac{dw}{a^2(w)} \mathcal{P}_\delta \left(\frac{\ell}{f_K(w)}, w \right) \times \left[\int_w^{w_h} dw' p_w(w') \frac{f_K(w' - w)}{f_K(w')} \right]^2, \quad (8)$$

with ℓ as the Fourier mode on the sky, w denotes the comoving coordinate, w_h the comoving coordinate of the horizon, $f_K(w)$ the comoving angular diameter distance and p_w the redshift distribution of source galaxies. The 2PCFs depend on both power spectra, \mathcal{P}_E and \mathcal{P}_B

$$\xi_+(\theta) = \int_0^\infty \frac{d\ell}{2\pi} \ell J_0(\ell\theta) [\mathcal{P}_E(\ell) + \mathcal{P}_B(\ell)], \quad (9)$$

$$\xi_-(\theta) = \int_0^\infty \frac{d\ell}{2\pi} \ell J_4(\ell\theta) [\mathcal{P}_E(\ell) - \mathcal{P}_B(\ell)], \quad (10)$$

with J_n denoting the n -th order Bessel function.

Another second-order cosmic shear measure, the aperture mass dispersion, was introduced by Schneider et al. (1998) and is also related to the power spectrum. In contrast to the 2PCF $\langle M_{\text{ap}}^2 \rangle$ only depends on the E-mode and $\langle M_\perp^2 \rangle$ only on the B-mode power spectrum, hence the aperture mass statistics provides a powerful tool to separate E- from B-modes

$$\langle M_{\text{ap}}^2 \rangle(\theta) = \frac{1}{2\pi} \int_0^\infty d\ell \ell \mathcal{P}_E(\ell) W_{\text{ap}}(\ell\theta), \quad (11)$$

$$\langle M_\perp^2 \rangle(\theta) = \frac{1}{2\pi} \int_0^\infty d\ell \ell \mathcal{P}_B(\ell) W_{\text{ap}}(\ell\theta), \quad (12)$$

with

$$W_{\text{ap}}(\theta\ell) = \left(\frac{24J_4(\ell\theta)}{(\ell\theta)^2} \right)^2. \quad (13)$$

From (9), (10) and (11) we see that the second-order shear measures are filtered versions of \mathcal{P}_E and \mathcal{P}_B . How the different filter functions influence the information content of the corresponding measures will be examined more closely in Sect. 3. In practice the aperture mass dispersion is difficult to measure due to gaps and holes in the data field but can be expressed in terms of ξ_+ and ξ_- as

$$\langle M_{\text{ap}}^2 \rangle(\theta) = \int_0^{2\theta} \frac{d\vartheta}{2\theta^2} \left[\xi_+(\vartheta) T_+ \left(\frac{\vartheta}{\theta} \right) + \xi_-(\vartheta) T_- \left(\frac{\vartheta}{\theta} \right) \right]. \quad (14)$$

The explicit calculation and the filter functions T_{\pm} are given in Schneider et al. (2002b).

2.2. Estimators

Consider a sample of galaxies with angular positions θ_i . For each pair of galaxies we define the connecting vector $\theta = \theta_i - \theta_j$ and determine tangential and cross-components of the ellipticities (ϵ_t and ϵ_x) with respect to this connecting vector. From these ellipticities we estimate the 2PCF in logarithmic bins of ϑ with a logarithmic bin width $\Delta\vartheta$ (Schneider et al. 2002a). If the bin width is sufficiently small an unbiased estimator for $\xi_{\pm}(\vartheta)$ is given by

$$\hat{\xi}_{\pm}(\vartheta) = \frac{1}{N_p(\vartheta)} \sum_{ij} (\epsilon_{it}\epsilon_{jt} \pm \epsilon_{ix}\epsilon_{jx}) \Delta\vartheta(|\theta_i - \theta_j|), \quad (15)$$

with $N_p(\vartheta) = \sum_{ij} \Delta\vartheta(|\theta_i - \theta_j|)$ as the number of galaxy pairs inside a bin and $\Delta\vartheta(|\theta_i - \theta_j|)$ is 1 if $|\theta_i - \theta_j|$ lies inside bin ϑ , 0 otherwise. An unbiased estimator of $\langle M_{\text{ap}}^2 \rangle$ can be calculated from $\hat{\xi}_{\pm}(\vartheta)$ using (14),

$$\mathcal{M}(\theta_k) = \sum_{i=1}^I \frac{\Delta\vartheta_i \vartheta_i}{2\theta_k^2} \left[\hat{\xi}_+(\vartheta_i) T_+ \left(\frac{\vartheta_i}{\theta_k} \right) + \hat{\xi}_-(\vartheta_i) T_- \left(\frac{\vartheta_i}{\theta_k} \right) \right], \quad (16)$$

where I must be chosen such that the upper limit of the I^{th} bin equals twice the value of θ_k .

2.3. Covariances

Important for characterizing the amount of information of a shear estimator is the corresponding covariance. For the 2PCF it is defined as

$$C_{\xi}(\vartheta_i, \vartheta_j) := \langle (\xi_{\pm}(\vartheta_i) - \hat{\xi}_{\pm}(\vartheta_i)) (\xi_{\pm}(\vartheta_j) - \hat{\xi}_{\pm}(\vartheta_j)) \rangle. \quad (17)$$

Assuming a Gaussian shear field the covariance of the 2PCF can be calculated analytically (Schneider et al. 2002a; Joachimi et al. 2007). As one already sees from (17) the 2PCF has four different covariances, denoted as C_{++} , C_{+-} , C_{-+} , C_{--} . Only three of them are independent since $C_{+-}(\vartheta_i, \vartheta_j) = C_{-+}(\vartheta_j, \vartheta_i)$. The covariance $C_{\mathcal{M}}(\theta_k, \theta_l)$ of \mathcal{M} is defined analogously. Using

(16) we can express $C_{\mathcal{M}}$ in terms of C_{ξ}

$$C_{\mathcal{M}}(\theta_k, \theta_l) = \frac{1}{4} \sum_{i=1}^I \sum_{j=1}^J \frac{\Delta\vartheta_i \Delta\vartheta_j}{\theta_k^2 \theta_l^2} \vartheta_i \vartheta_j \times \left[\sum_{m,n=+,-} T_m \left(\frac{\vartheta_i}{\theta_k} \right) T_n \left(\frac{\vartheta_j}{\theta_l} \right) C_{mn}(\vartheta_i, \vartheta_j) \right]. \quad (18)$$

Similar to (16) I (J) are chosen such that the upper limit of the I^{th} (J^{th}) bin equals twice of θ_k (θ_l).

3. The new data vector \mathcal{N}

Consider two data vectors, namely

$$\xi = \begin{pmatrix} \xi_+ \\ \xi_- \end{pmatrix} \quad \text{with} \quad \xi_+ = \begin{pmatrix} \xi_+(\vartheta_1) \\ \vdots \\ \xi_+(\vartheta_m) \end{pmatrix}, \quad \xi_- = \begin{pmatrix} \xi_-(\vartheta_1) \\ \vdots \\ \xi_-(\vartheta_m) \end{pmatrix} \quad (19)$$

for the 2PCF and

$$\langle M_{\text{ap}}^2 \rangle = \begin{pmatrix} \langle M_{\text{ap}}^2 \rangle(\theta_1) \\ \vdots \\ \langle M_{\text{ap}}^2 \rangle(\theta_n) \end{pmatrix} \quad (20)$$

for the aperture mass dispersion. The relation (16) can also be written in terms of data vectors and a $n \times 2m$ transfer matrix \mathbf{A}

$$\langle M_{\text{ap}}^2 \rangle = \underbrace{(\mathbf{A}_+ | \mathbf{A}_-)}_{\mathbf{A}} \begin{pmatrix} \xi_+ \\ \xi_- \end{pmatrix}, \quad (21)$$

with \mathbf{A}_+ denoting the part of \mathbf{A} referring to ξ_+ and \mathbf{A}_- denotes the corresponding part referring to ξ_- . Eq. (21) implies that the information content of $\langle M_{\text{ap}}^2 \rangle$ is less or equal compared to ξ . The amount of information can only be equal if and only if the rank of \mathbf{A} equals the dimension of ξ , hence $\text{rank } \mathbf{A} = 2m$. We explicitly prove these statements in the Appendix. For the case of ξ and $\langle M_{\text{ap}}^2 \rangle$ $n \leq m$ holds, which can be seen from (16). Therefore the relation (21) is not invertible and the information content of $\langle M_{\text{ap}}^2 \rangle$ is smaller compared to ξ_{\pm} . The fact that ξ_{\pm} contains more information on cosmological parameters can also be explained when looking at the filter functions J_0 , J_4 and W_{ap} relating the corresponding second-order shear measures to the underlying power spectrum. ξ_+ probes the power spectrum over a broad range of Fourier modes and collects information also on scales larger than the survey size. In contrast, the aperture mass dispersion provides a highly localized probe of \mathcal{P}_E and does not contain this large-scale information. Hence, due to the limited field size of a survey the information content of $\langle M_{\text{ap}}^2 \rangle$ is smaller compared to ξ_{\pm} . These considerations lead to the idea to modify $\langle M_{\text{ap}}^2 \rangle$ by adding one data point of $\xi_+(\theta_0)$. We define the new data vector \mathcal{N} as

$$\mathcal{N} = \begin{pmatrix} \langle M_{\text{ap}}^2 \rangle(\theta_1) \\ \vdots \\ \langle M_{\text{ap}}^2 \rangle(\theta_n) \\ \xi_+(\theta_0) \end{pmatrix} \quad (22)$$

and the corresponding covariance matrix reads

$$\mathbf{C}_N = \begin{pmatrix} \mathbf{C}_{\mathcal{M}_{11}} & \cdots & \mathbf{C}_{\mathcal{M}_{1n}} & | & \mathbf{C}(\mathcal{M}_1, \xi_+) \\ \vdots & \ddots & \vdots & | & \vdots \\ \mathbf{C}_{\mathcal{M}_{n1}} & \cdots & \mathbf{C}_{\mathcal{M}_{nn}} & | & \mathbf{C}(\mathcal{M}_n, \xi_+) \\ \hline \mathbf{C}(\xi_+, \mathcal{M}_1) & \cdots & \mathbf{C}(\xi_+, \mathcal{M}_n) & | & \mathbf{C}(\xi_+, \xi_+) \end{pmatrix}. \quad (23)$$

The upper left $n \times n$ matrix is exactly $\mathbf{C}_{\mathcal{M}}$ and the entry for $\mathbf{C}(\xi_+, \xi_+)$ is taken from the corresponding covariance matrix of the correlation function. The cross terms can be calculated using (16) and read

$$\begin{aligned} \mathbf{C}(\mathcal{M}(\theta_k), \hat{\xi}_+(\theta_0)) = & \frac{1}{2} \sum_{i=1}^I \frac{\Delta\vartheta_i}{\theta_k^2} \vartheta_i \left[T_+ \left(\frac{\vartheta_i}{\theta_k} \right) \mathbf{C}_{++}(\vartheta_i, \theta_0) \right. \\ & \left. + T_- \left(\frac{\vartheta_i}{\theta_k} \right) \mathbf{C}_{-+}(\vartheta_i, \theta_0) \right]. \end{aligned} \quad (24)$$

4. Calculating data vectors and covariances

The data vectors $\xi, \langle M_{\text{ap}}^2 \rangle, \mathcal{N}$ are directly calculated from the power spectrum of density fluctuations \mathcal{P}_δ using (8) to obtain \mathcal{P}_E and then applying either (9), (10) or (11) depending on the desired cosmic shear measure. To derive \mathcal{P}_δ we assume an initial Harrison-Zeldovich power spectrum ($\mathcal{P}_\delta(k) \propto k^n$ with $n = 1$). The transition to todays power spectrum employs the transfer function described in Bardeen et al. (1986), and for the calculation of the nonlinear evolution we use the fitting formula of Smith et al. (2003). In contrast, the covariances are obtained from ray-tracing simulations. The N-body simulation used for the ray-tracing experiment was carried out by the Virgo Consortium (Jenkins et al. 2001); for details of the ray-tracing algorithm see Ménard et al. (2003). \mathbf{C}_ξ is calculated by field-to-field variation of 36 ray-tracing realisations, where each field has a sidelength of 4.27 degrees. The intrinsic ellipticity noise is $\sigma_\epsilon = 0.3$ and the number density of source galaxies is given by $n = 25/\text{arcmin}^2$. From \mathbf{C}_ξ we calculate $\mathbf{C}_{\mathcal{M}}$ and \mathbf{C}_N according to (18) and (23). The cosmology of the ray-tracing simulations, i.e. our fiducial cosmological model is a flat Λ CDM model with $\Omega_m = 0.3$, $\sigma_8 = 0.9$, $h = 0.7$ and $\Gamma = 0.172$.

4.1. Difficulties with covariances

4.1.1. Underestimation of $\mathbf{C}_{\mathcal{M}}$

Kilbinger et al. (2006) have shown that $\langle M_{\text{ap}}^2 \rangle(\theta)$ is biased for small θ when calculated from the 2PCF using (16). This is due to the lack of 2PCF data points on very small angular scales which causes a small-scale cutoff in the integral of (14). In our specific case the $\langle M_{\text{ap}}^2 \rangle$ data vector is not affected by this bias because we calculate it directly from the power spectrum \mathcal{P}_E . However, since $\mathbf{C}_{\mathcal{M}}$ and \mathbf{C}_N are calculated from the covariance of the 2PCF, they are certainly affected by this problem. In this subsection we determine the θ -range on which we can calculate $\mathbf{C}_{\mathcal{M}}$ with sufficient accuracy; the corresponding data vector of the aperture mass dispersion will be restricted to this range. Fig. 1 shows $\langle M_{\text{ap}}^2 \rangle$ calculated directly from the power spectrum compared with $\langle M_{\text{ap}}^2 \rangle$ calculated from ξ_\pm using (16). We

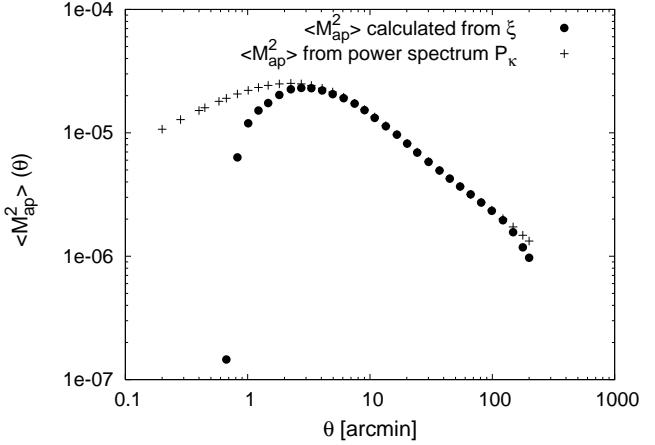


Fig. 1. This plot shows $\langle M_{\text{ap}}^2 \rangle$ calculated directly from the power spectrum (11) compared to $\langle M_{\text{ap}}^2 \rangle$ calculated from ξ_\pm (16). Due to the fact that we cannot estimate the 2PCF down to arbitrary small angular scales (here $\vartheta_{\text{min}} = 0.2$) the calculated $\langle M_{\text{ap}}^2 \rangle$ values are underestimated. The same problem occurs when calculating $\mathbf{C}_{\mathcal{M}}$ from \mathbf{C}_ξ . The θ -range with a deviation smaller than 5% last from 2°25 – 100°0.

assume that the deviation shown here is a good approximation for the bias in $\mathbf{C}_{\mathcal{M}}$ and we require an accuracy of 5 % to accept a θ -value for the $\langle M_{\text{ap}}^2 \rangle$ data vector. This criterion restricts the data vector to a θ -range of 2°25 – 100°0 whereas the 2PCF data vector is measured from 0°2 – 200°0.

4.1.2. Inversion of the covariance matrix

A second difficulty in the context of covariance matrices is outlined in Hartlap et al. (2007). The fact that an inversion of an estimated unbiased covariance matrix leads to a biased result can be overcome by applying a correction factor. According to Hartlap et al. (2007) the correction factor depends on the ratio of number of bins (B) to number of independent realisations (N) from which the covariance matrix is estimated. An unbiased estimate of the inverse covariance matrix is

$$\mathbf{C}_{\text{unbiased}}^{-1} = \frac{N - B - 2}{N - 1} \mathbf{C}^{-1} = \left[1 - \frac{B + 1}{N - 1} \right] \mathbf{C}^{-1}. \quad (25)$$

Hartlap et al. (2007) have proven the validity of this correction factor for the case of Gaussian errors and statistically independent data vectors. These two assumptions are violated when estimating the covariance matrix from ray-tracing simulations. In order to check whether the correction factor corrects the error in our ray-tracing covariance matrices, we perform the following experiment. We add different Gaussian noise to the ellipticities of the galaxies, which are taken from the 36 independent realisations of the ray-tracing simulations and thereby increase the number of independent realisations. We hold the binning of the matrices constant, calculate covariances for 36, 108, 216, 360, 720, 1080, 1440, 1800 independent realisations and plot $1/\text{tr } \mathbf{C}^{-1}$ depending on the ratio B/N (Fig. 2). Note that this method only creates multiple realisations of Gaussian noise on the galaxy ellipticities and does not increase the number of realisations which determine the cosmic variance part of the covariance matrix. Therefore, this method only partly checks for

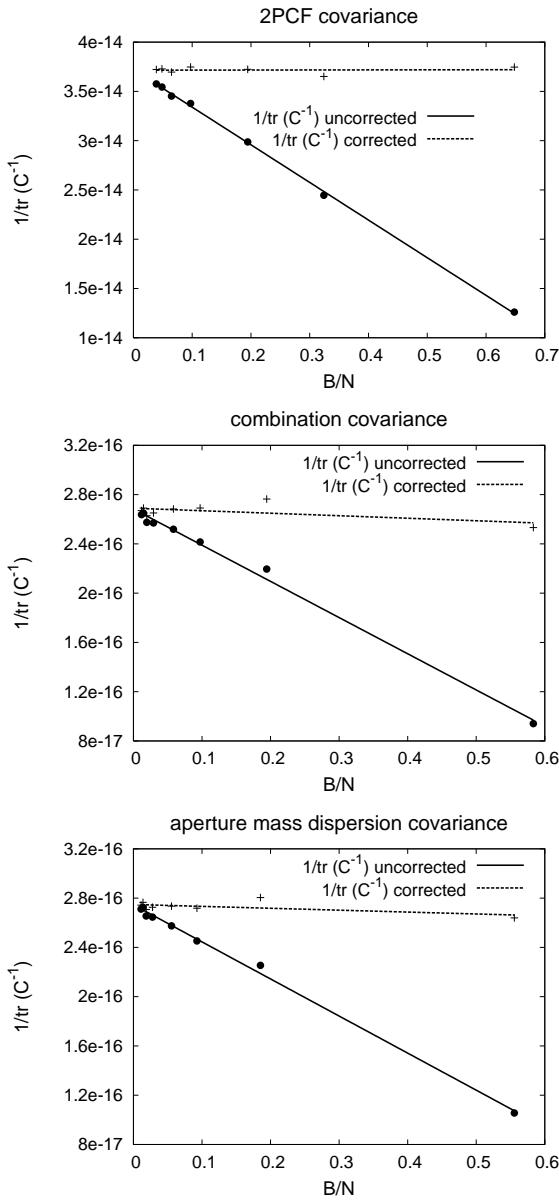


Fig. 2. Inverting an estimated covariance matrix yields a bias which depends on the ratio of the number of bins to the number of independent realisations of the ray-tracing simulations (B/N). This dependence is linear and we correct the bias for all three inverted covariance matrices \mathbf{C}_ξ , \mathbf{C}_N , \mathbf{C}_M . We plot $1/\text{tr}(\mathbf{C}^{-1})$ for the corrected and uncorrected values where the lines indicate a fit through the data. All covariances are binned logarithmically, \mathbf{C}_ξ consists of 70 bins covering a range $0'2 - 200'0$; \mathbf{C}_N and \mathbf{C}_M cover the range $2'25 - 100'0$ with 21 bins for \mathbf{C}_N and 20 bins for \mathbf{C}_M .

the non-Gaussianity of the errors in a ray-tracing covariance matrix, nevertheless the impact of statistically dependent data vectors is fully taken into account. We find the same linear behavior of the bias as Hartlap et al. (2007), therefore we are confident that the correction factor is able to unbias our covariance matrices. Using the corrected inverse covariance matrix we assure that the log-likelihood is also unbiased, nevertheless, any non-linear transformation of the log-likelihood will again introduce a bias which influences the results and must be examined.

5. Likelihood analysis

We define the posterior likelihood (P_{PL}) for the case of a 2PCF data vector as

$$P_{PL}(\boldsymbol{\xi}|\boldsymbol{\pi}) = \frac{P_L(\boldsymbol{\xi}|\boldsymbol{\pi})}{P_E(\boldsymbol{\xi})} P_{\text{Prior}}(\boldsymbol{\pi}), \quad (26)$$

where $\boldsymbol{\pi}$ denotes the parameter vector of the Λ CDM model assumed in our likelihood analysis. P_{Prior} usually contains knowledge on the parameter vector from other experiments. In our case we assume flat priors with cutoffs, which means P_{Prior} is constant for all parameters inside a fixed interval and $P_{\text{Prior}} = 0$ for parameters outside the interval. The evidence P_E , is just the normalization, obtained by integrating the probability over the whole parameter space. The likelihood P_L , is defined as

$$P(\boldsymbol{\xi}|\boldsymbol{\pi}) = \frac{1}{(2\pi)^{n/2} \sqrt{\det \mathbf{C}_\xi}} \exp \left[-\frac{1}{2} \chi^2(\boldsymbol{\xi}, \boldsymbol{\pi}) \right], \quad (27)$$

with the χ^2 -function

$$\chi^2(\boldsymbol{\xi}, \boldsymbol{\pi}) = (\boldsymbol{\xi}(\boldsymbol{\pi}) - \boldsymbol{\xi}^f)^t \mathbf{C}_\xi^{-1} (\boldsymbol{\xi}(\boldsymbol{\pi}) - \boldsymbol{\xi}^f). \quad (28)$$

$\boldsymbol{\xi}^f$ denotes the data vector corresponding to our fiducial model, whereas $\boldsymbol{\xi}(\boldsymbol{\pi})$ varies according to the considered parameter space. To compare the information content of $\boldsymbol{\xi}$, $\langle M_{\text{ap}}^2 \rangle$, \mathcal{N} we calculate the posterior likelihood in several parameter spaces and illustrate the result by contour plots. Smaller contours correspond to a higher information content.

5.1. Quadrupole moments

In addition to contour plots we illustrate the information content of a data vector by calculating the determinant of the quadrupole moment of the posterior likelihood (Kilbinger & Schneider 2004)

$$Q_{ij} \equiv \int d^2\boldsymbol{\pi} P_{PL}(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2) (\boldsymbol{\pi}_i - \boldsymbol{\pi}_i^f)(\boldsymbol{\pi}_j - \boldsymbol{\pi}_j^f), \quad (29)$$

with $\boldsymbol{\pi}_1$ and $\boldsymbol{\pi}_2$ as the varied parameters, $\boldsymbol{\pi}_i^f$ as the parameter of the fiducial model. The calculation of Q_{ij} assumes a posterior likelihood in a two-dimensional parameter space, when considering more than two varied parameters we calculate the Q_{ij} for the marginalized posterior likelihood (see Sect. 5.3). The determinant is given by

$$q = \sqrt{\det Q_{ij}} = \sqrt{Q_{11}Q_{22} - Q_{12}^2}. \quad (30)$$

Tighter constraints on the parameters correspond to a smaller value of q . Due to its non-linearity in the log-likelihood q is biased (Sect. 4.1.2). The amount of bias varies depending on the number of independent realisations from which the covariance matrix is estimated and we examine this effect in a similar way as for the covariance matrices in Sect. 4.1.2. For six different numbers of independent realisations we perform a likelihood analysis in a two-parameter space (Ω_m vs. σ_8) and calculate q for all three cosmic shear measures. The result is plotted in Fig. 3. One clearly sees that the q dependence on the number of realisations is much weaker compared with the difference between q of different cosmic shear measures. Therefore, the bias is small and we can confidently use q to compare the relative information content of the different data vectors.

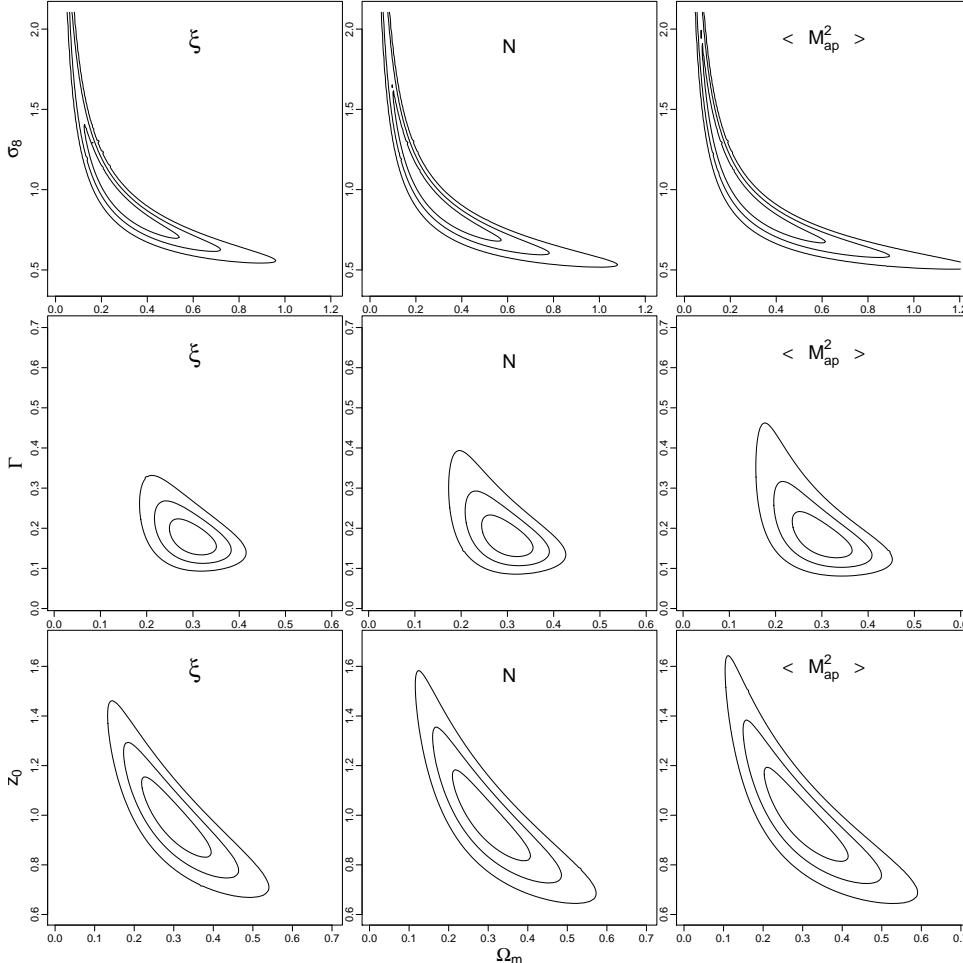


Fig. 5. The likelihood contours when varying only two parameters, while the others are fixed to the fiducial values. The contours contain 68.3 %, 95.4 %, 99.73 % of the posterior likelihood. We consider 3 parameter spaces, from top to bottom: σ_8 vs. Ω_m , Γ vs. Ω_m , z_0 vs. Ω_m . The constraints of ξ are shown on the left, \mathcal{N} is plotted in the middle and the results of $\langle M_{ap}^2 \rangle$ are shown on the right.

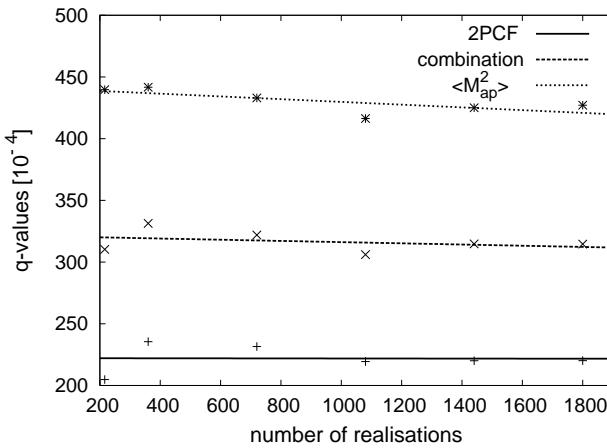


Fig. 3. The q of $\xi, \mathcal{N}, \langle M_{ap}^2 \rangle$ depending on the numbers of independent realisations from which the covariance matrix is estimated. As a parameter space we chose Ω_m vs. σ_8 . The deviation of q belonging to different numbers of realisations is much smaller than the difference of q of different data vectors.

5.2. Variations of two parameters

The likelihood analysis in this section is performed in a two-dimensional parameter space; all other cosmological parameters are fixed to the fiducial values. Before comparing the three

data vectors we optimize \mathcal{N} with respect to the θ_0 -value of the 2PCF data point. We add 35 different $\xi_+(\theta_0)$ covering a range $\theta_0 \in [0.2^\circ, 200.0^\circ]$ and calculate q . Fig. 4 illustrates the results of this optimization for 3 different pairs of parameters (Γ vs. Ω_m , σ_8 vs. Ω_m , z_0 vs. Ω_m). For all parameter combinations considered the optimal θ_0 is close to 10° . This can be explained from the behavior of the covariance matrix. For small angular scales the covariance is dominated by shot noise, whereas for large angular scales the signal of ξ_+ becomes very small. In both cases the signal-to-noise ratio is lower than at medium angular scales, where we find the minimum of q . In our further analysis we always choose the optimal 2PCF data point for the combined data vector. The results are illustrated by contour plots (Fig. 5) and the corresponding values of q are summarized in Table 1. Here, we also list the results for two additional parameter combinations, σ_8 vs. Γ and z_0 vs. σ_8 , not shown in Fig. 5. One clearly sees that the 2PCF data vector gives the tightest constraints on cosmological parameters whereas constraints from the aperture mass dispersion are weaker. Although not matching quite the amount of information of ξ , the combined data vector is a substantial improvement compared to $\langle M_{ap}^2 \rangle$. This result is consistent for all parameter combinations we examine; nevertheless the amount of the improvement varies. We calculate the difference in information of ξ and \mathcal{N} relative to $\langle M_{ap}^2 \rangle$ and denote these values

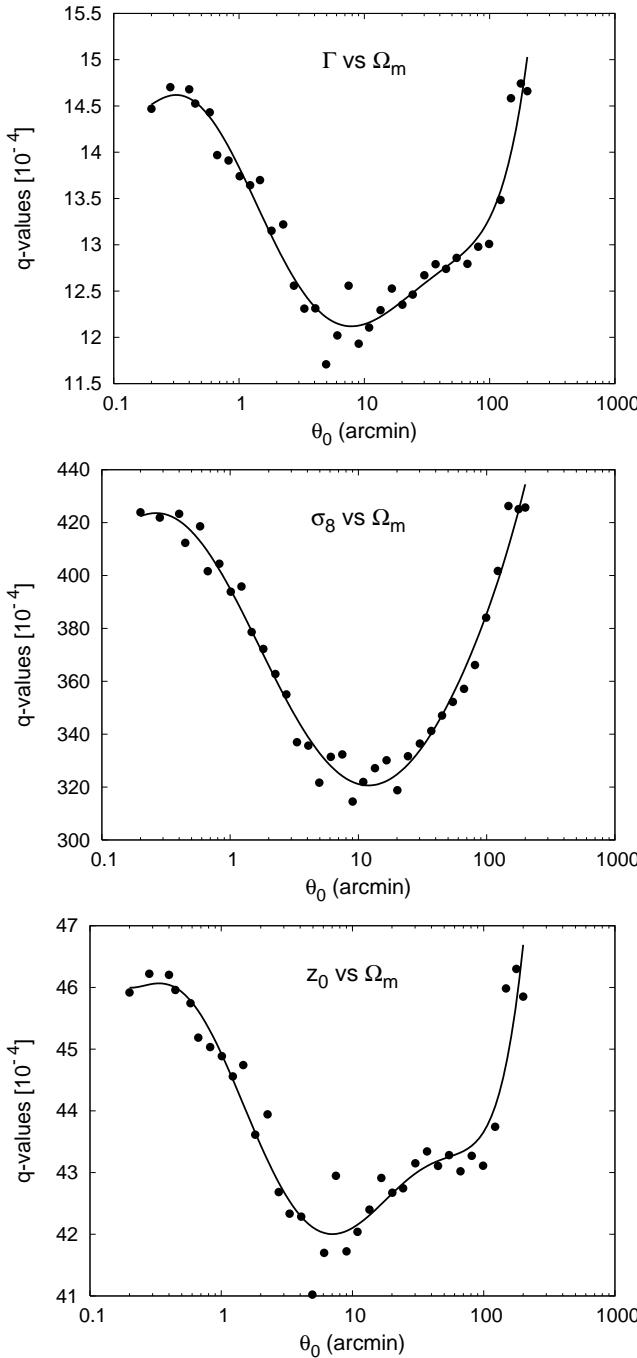


Fig. 4. Here we plot q of the combined data vector when varying (θ_0) of the additional ξ_+ data point. We calculate q for 35 different added $\xi_+(\theta_0)$ and show the behaviour in three different parameter spaces. q of the combined data vector can be optimized with respect to θ_0 and the optimal values are 7.8 (Γ vs. Ω_m), 12.9 (σ_8 vs. Ω_m) and 7.0 (z_0 vs. Ω_m). These values are the minima of a polynomial fit through the data points.

$\Delta\xi$ and $\Delta\mathcal{N}$ (Table 1). The parameter combination σ_8 vs. Ω_m shows a relative improvement of $\Delta\mathcal{N} = 26.4\%$, whereas the improvement is much less for the case z_0 vs. σ_8 ($\Delta\mathcal{N} = 4.1\%$). The amount of new information of $\xi_+(\theta_0)$ depends on two main issues. First, ξ_+ integrates over a very broad range of the power spectrum and it can happen that although \mathcal{P}_E is sensitive to the

Table 1. This table shows the q of ξ , \mathcal{N} and $\langle M_{ap}^2 \rangle$ for various parameter spaces. Parameters over which we marginalize are mentioned in brackets. The entries are given in units of 10^{-4} and only q of the same parameter space can be compared. $\Delta\mathcal{N}$ ($\Delta\xi$) gives the relative improvement compared to the q of $\langle M_{ap}^2 \rangle$ and this improvement differs with respect to the parameter space.

parameter space	$\langle M_{ap}^2 \rangle$	\mathcal{N}	ξ	$\Delta\mathcal{N}$	$\Delta\xi$
Γ vs. Ω_m	14.7	11.7	9.1	20.4 %	38.1 %
σ_8 vs. Γ	23.1	19.0	14.6	17.8 %	36.8 %
σ_8 vs. Ω_m	427.1	314.5	220.1	26.4 %	48.5 %
z_0 vs. Ω_m	46.4	41.0	32.9	11.6 %	29.1 %
z_0 vs. σ_8	95.3	91.4	73.2	4.1 %	23.2 %
σ_8 vs. Ω_m (z_0)	416.9	313.4	230.0	25.8 %	44.8 %
σ_8 vs. Ω_m (Γ)	780.5	720.9	527.0	7.6 %	32.5 %
Γ vs. Ω_m (σ_8)	93.7	77.6	61.6	17.2 %	34.3 %
σ_8 vs. Ω_m (Γ, z_0)	983.8	850.6	623.5	13.5 %	36.6 %

parameters considered, the integral over \mathcal{P}_κ is much less. For example, if one varies Γ , the power spectrum is tilted and looks significantly different, whereas the corresponding $\xi_+(\theta_0)$ might be very similar. Second, $\langle M_{ap}^2 \rangle$ does not contain information on small Fourier modes, whereas \mathcal{N} gains information about these modes from the data point $\xi_+(\theta_0)$. However, in case these modes of the power spectrum are not sensitive to parameters considered, the information which is contributed by $\xi_+(\theta_0)$ is mainly redundant, hence $\Delta\mathcal{N}$ is low. For example, varying σ_8 or Ω_m changes \mathcal{P}_E similarly, i.e. increasing Ω_m or σ_8 increases the amplitude of \mathcal{P}_E on all Fourier modes. Therefore, the integration over \mathcal{P}_E is equally sensitive to parameter variations as \mathcal{P}_E itself. Furthermore the deviation of power spectra with different values in σ_8 and Ω_m becomes much more significant for small Fourier modes. Information on these scales is not included in $\langle M_{ap}^2 \rangle$ but contributed by $\xi_+(\theta_0)$, resulting in a large $\Delta\mathcal{N}(26.4\%)$. In contrast to this, a variation of z_0 changes the power spectrum very little, especially on low ℓ -scales the dependence is weak. Accordingly, the gain in information for the cases z_0 vs. Ω_m and z_0 vs. σ_8 is rather small.

5.3. Variation of three and four parameters - marginalization

In this section we perform a likelihood analysis in three- and four-dimensional parameter space. To illustrate the results in two-dimensional contour plots we define the marginalized posterior likelihood

$$P_{mPL}(\boldsymbol{\pi}_{12}|\xi_\pm) = \int d\boldsymbol{\pi}_3 \int d\boldsymbol{\pi}_4 P_{PL}(\boldsymbol{\pi}_{1234}|\xi_\pm), \quad (31)$$

which is obtained by integrating over the posterior likelihood of the marginalized parameters. The marginalized likelihood is also biased due to its non-linearity in the log-likelihood. To examine whether this bias affects our results significantly we perform the same experiment as done for q in two-dimensional parameter space. We calculate q for our three different measures depending on the number of realisations. The results are

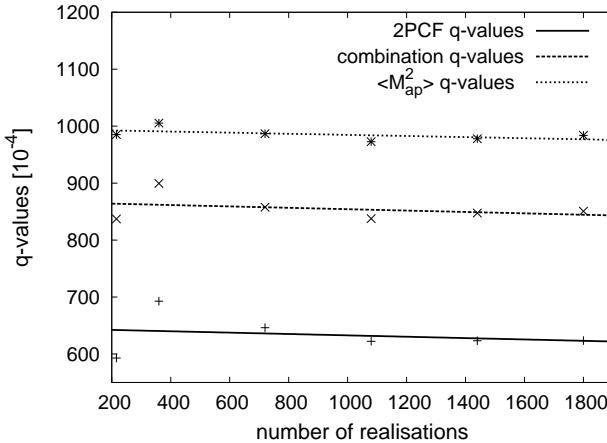
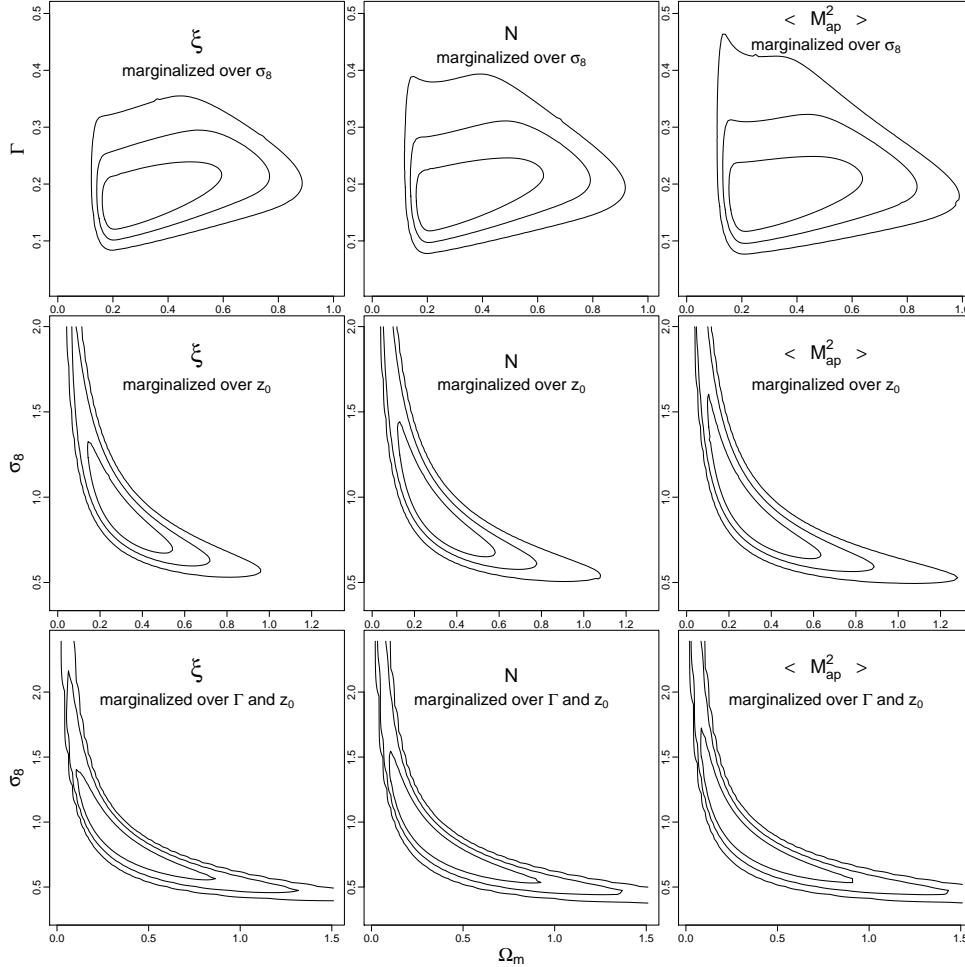


Fig. 6. This figure shows the q of ξ , \mathcal{N} and $\langle M_{ap}^2 \rangle$ for the marginalized posterior likelihood depending on different numbers of realisations. The parameter space is σ_8 vs. Ω_m (marginalized over Γ and z_0). The deviation of q belonging to different numbers of realisations is much smaller compared to the deviation of q of different measures. The lines indicate the fit through the data points.

shown in Fig. 6; again, the bias due to the process of marginalization is small compared to the difference of q of our three data vectors showing that also in the marginalized case we can

Fig. 7. The likelihood contours of ξ , \mathcal{N} and $\langle M_{ap}^2 \rangle$ in three- and four-dimensional parameter space. From top to bottom we see Γ vs. Ω_m marginalized over σ_8 , σ_8 vs. Ω_m marginalized over z_0 and σ_8 vs. Ω_m marginalized over Γ and z_0 . The contours contain 68.3 %, 95.4 %, 99.73 % of the marginalized posterior likelihood. The small scatter of the contours in the last plot is due to a lower resolution of the grid in four-dimensional parameter space compared to the grids in two- and three-dimensional parameter space. The contours, although broader, are comparable to those given in Fig. 5.

Table 2. The optimal angular separation θ_0 for the added ξ_+ in the combined data vector \mathcal{N} . The values are comparable to the similar analysis in two-parameter space (see Fig. 4).

parameter space	optimal value θ_0
Γ vs. Ω_m (marginalized over σ_8)	$\theta_0 = 9'.1$
σ_8 vs. Ω_m (marginalized over z_0)	$\theta_0 = 13'.0$
σ_8 vs. Ω_m (marginalized over Γ and z_0)	$\theta_0 = 12'.0$

use q to compare the information content. We also optimize the combined data vector, similar to Sect. 5.2 and summarize the results in Table 2. For the same reasons as in the previous section the optimal angular scale of the added ξ_+ data point is again around $10'$ and we choose this optimized \mathcal{N} for the likelihood analysis in three- and four-dimensional parameter space. The results of the likelihood analysis are comparable to those obtained in two-dimensional parameter space. The q (see Table 1) are larger and the contours (see Fig. 7) are broader. Again, the relative improvement $\Delta\mathcal{N}$ depends on the parameter space considered. For σ_8 vs. Ω_m marginalized over z_0 the improvement is very high (25.8 %) but becomes much lower for σ_8 vs. Ω_m marginalized over Γ . This can be explained by looking how \mathcal{P}_E changes with respect to the variation in parameter space. For the combination σ_8 vs. Ω_m , we already explained

this in Sect. 5.2 and the influence of z_0 on \mathcal{P}_E is quite similar. Increasing z_0 also increases \mathcal{P}_E , although the effect is not very large. Therefore, the improvement of σ_8 vs. Ω_m marginalized over z_0 is comparable to the non-marginalized case. When varying the shape parameter Γ , \mathcal{P}_E is tilted and this dependence of \mathcal{P}_E on Γ is different compared to the other three parameters. Scales of \mathcal{P}_E which are most sensitive to Γ differ from scales sensitive to σ_8 , Ω_m and z_0 and the same argument holds for the scales of the added $\xi_+(\theta_0)$. Therefore, the optimal θ_0 for the case σ_8 vs. Ω_m marginalized over Γ is a compromise and the relative improvement is much lower (7.6 %) compared to σ_8 vs. Ω_m marginalized over z_0 (25.8 %).

6. Simulation of a B-mode contamination on small angular scales

In this section we simulate a B-mode contamination of ξ , \mathcal{N} and $\langle M_{ap}^2 \rangle$ on small angular scales. At present there is no model available which describes B-modes; taking into account that B-modes most likely occur on small angular scales (e.g. Hoekstra et al. 2002; van Waerbeke et al. 2005; Massey et al. 2007) we use the following arbitrary model for a B-mode power spectrum

$$\mathcal{P}_B(\ell) = 0.2 \mathcal{P}_E(\ell) e^{-\ell_B/\ell}, \quad (32)$$

where ℓ_B defines a scale beyond which the B-mode contamination decreases quickly. The B-mode contribution to ξ can be calculated from (9) and (10) by assuming $\mathcal{P}_E = 0$. In order to calculate the covariance \mathbf{C}_B we assume that the probability distribution of B-modes can be described by a Gaussian random field. This assumption enables us to calculate the covariance directly in terms of the power spectrum \mathcal{P}_B (Joachimi et al. 2007). The covariance of the 2PCF corresponding to the B-mode contribution is given by

$$\begin{aligned} C_{B,ij}^{++} &= \frac{1}{A\pi} \int d\ell \ell J_0(\ell\vartheta_i) J_0(\ell\vartheta_j) \left(\mathcal{P}_B^2(\ell) + \mathcal{P}_B(\ell) \frac{\sigma_\epsilon^2}{n} \right), \\ C_{B,ij}^{--} &= \frac{1}{A\pi} \int d\ell \ell J_4(\ell\vartheta_i) J_4(\ell\vartheta_j) \left(\mathcal{P}_B^2(\ell) + \mathcal{P}_B(\ell) \frac{\sigma_\epsilon^2}{n} \right), \\ C_{B,ij}^{+-} &= -\frac{1}{A\pi} \int d\ell \ell J_0(\ell\vartheta_i) J_4(\ell\vartheta_j) \left(\mathcal{P}_B^2(\ell) + \mathcal{P}_B(\ell) \frac{\sigma_\epsilon^2}{n} \right), \end{aligned}$$

where A defines the volume of the survey, σ_ϵ the intrinsic ellipticity noise and n the number density of the source galaxies. According to the corresponding values of the ray-tracing simulations we choose $\sigma_\epsilon = 0.3$ and $n = 25/\text{arcmin}^2$. Note that $C_{B,ij}^{-+} = C_{B,ji}^{+-}$. The pure shot noise term of $\mathbf{C}_B^{\pm\pm}$ is contained in $\mathbf{C}_E^{\pm\pm}$, in case of \mathbf{C}_B^{+-} this term vanishes anyway. We further assume that the contamination is independent of the lensing signal, hence there is no correlation between E- and B-modes. This assumption does not hold in case the B-mode signal is caused by insufficient PSF correction or other systematics, and we will comment on this at the end of this section. For the case that B-modes are created independently from E-modes we can define a combined E/B-mode covariance matrix as

$$\mathbf{C}_{\text{tot}} = \mathbf{C}_E + \mathbf{C}_B. \quad (33)$$

Recall that \mathbf{C}_E is estimated from ray-tracing simulations whereas \mathbf{C}_B is calculated by assuming a Gaussian random field. The correction factor, needed to invert estimated matrices correctly (see Sect. 4.1.1), must only be applied to \mathbf{C}_E , not for \mathbf{C}_B . We use the iterative approach of Miller (1981) to decompose this inverse of a sum of matrices into a summation of inverse matrices and then apply the correction factor only to \mathbf{C}_E^{-1} . From now on, the procedure of the comparison is similar to Sect. 5.2 and Sect. 5.3. We calculate the \mathbf{C}_M and \mathbf{C}_N from \mathbf{C}_ξ and perform a likelihood analysis. We only show the results for the Ω_m vs. σ_8 plane (see Fig. 8). The black dots indicate the fiducial cosmological model, and in case of the 2PCF data vector there is a significant deviation to the parameters of the maximum of the posterior likelihood. $\langle M_{ap}^2 \rangle$ and \mathcal{N} are much more robust against the contamination. As expected, the maximum of the posterior likelihood of the aperture mass dispersion matches exactly the fiducial parameters and in case of \mathcal{N} the discrepancy is negligibly small. Furthermore, the combination still gives tighter constraints on the parameters. As already mentioned above, the assumption of B-modes being independent of the E-mode signal does not always hold. In case the contamination affects both, E-mode and B-mode signal, the impact on the parameter constraints of the different measures is hard to quantify. In case one measures a B-modes signal, it is a common approach to assume that the E-mode signal is contaminated in a similar way, hence one correspondingly increases its error bars. Although this assumption is sensible, there are possible scenarios where the amount of contamination in E- and B-mode differs and the E-mode contamination cannot be quantified at all. Under the assumption that B-modes trace the scales of the E-mode contamination it is reasonable to exclude those scales from the likelihood analysis. This can be done using $\langle M_{ap}^2 \rangle$ or \mathcal{N} but ξ cannot avoid the contamination due to its broad filter functions.

7. Conclusions

Although the 2PCF and the aperture mass dispersion are both filtered versions of the power spectrum the first contains more information on \mathcal{P}_E than the latter. Reason for this is that ξ samples the power spectrum over a much broader range and also collects information on scales which are larger than the size of the survey. $\langle M_{ap}^2 \rangle$ lacks this large-scale information, but yields highly localized information on \mathcal{P}_E . Nevertheless $\langle M_{ap}^2 \rangle$ has other advantages. First, due to its narrow filter function the data points are much less correlated compared to the 2PCF data points. This leads to a mainly diagonal covariance matrix, which is numerically more stable during the inversion process in a likelihood analysis. Second, when considering higher-order statistics $\langle M_{ap}^3 \rangle$ is much easier to handle than the three-point correlation function (Schneider et al. 2005) and third, the aperture mass dispersion is sensitive to E-modes only. Based on these considerations we create the combined data vector \mathcal{N} , which preserves the advantages of $\langle M_{ap}^2 \rangle$ and additionally provides large-scale information on \mathcal{P}_E . This data vector can be optimized with respect to the angular scale of the added data point $\xi_+(\theta_0)$, but this optimization very likely depends on the survey geometry and must be performed for each

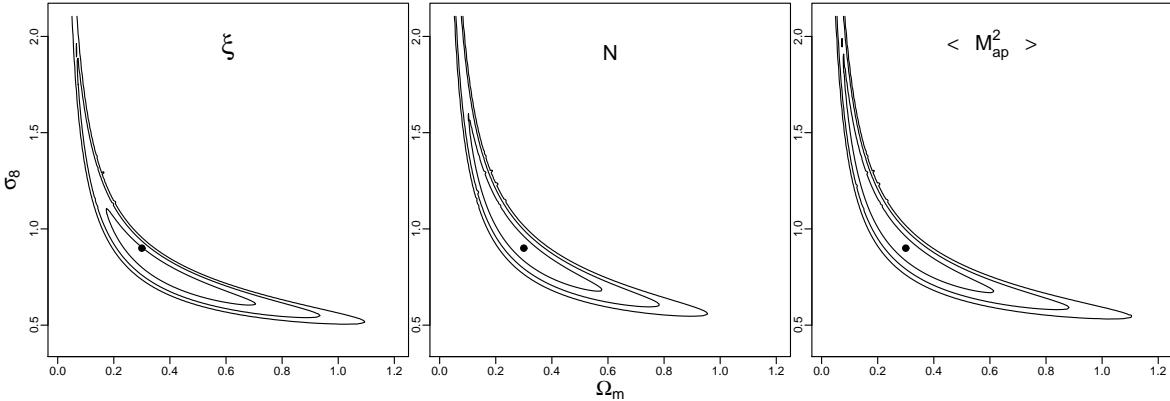


Fig. 8. This plot shows the likelihood contours for the case that the shear signal is contaminated with B-modes. We only consider a two-dimensional parameter space (σ_8 vs. Ω_m) and the contours again contain 68.3 %, 95.4 %, 99.73 % of the posterior likelihood. The black dot in each plot indicates the fiducial model. ξ gives biased constraints, \mathcal{N} and $\langle M_{\text{ap}}^2 \rangle$ are hardly contaminated.

survey separately. We compare the three data vectors in a detailed likelihood analysis and find that the combined data vector is a strong improvement in information content compared to $\langle M_{\text{ap}}^2 \rangle$. However, the amount of improvement depends on the parameter space considered, more precisely, on the dependence of \mathcal{P}_E on variation of those parameters. The combined data vector \mathcal{N} also maintains the other advantages of the aperture mass dispersion. Its covariance matrix is almost diagonal, even the cross terms $C(\mathcal{M}(\theta_k), \hat{\xi}_+(\theta_0))$ are much smaller compared with the off-diagonal terms of C_ξ . Comparing the information content of ξ and \mathcal{N} , ξ gives tighter constraints if the shear signal only consists of E-modes. In the more realistic case, when also B-modes are present, ξ is biased whereas \mathcal{N} is hardly affected and still gives tighter constraints on cosmological parameters compared to $\langle M_{\text{ap}}^2 \rangle$.

Appendix A: Comparison of two measures

We compare the information content of two arbitrary data vectors referring to them as primary data vector \mathbf{p} and secondary data vector \mathbf{s} . We further assume that \mathbf{s} can be calculated from \mathbf{p} by a *transfer matrix* \mathbf{A} (dimension $n \times m$), with arbitrary n and m

$$\mathbf{p} = \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_m \end{pmatrix} \quad \text{and} \quad \mathbf{s} = \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{pmatrix} \quad \text{with} \quad \mathbf{s} = \mathbf{A} \mathbf{p}. \quad (\text{A.1})$$

We define the covariance matrices of these data vectors as

$$\mathbf{C}_p = \langle (\mathbf{p} - \hat{\mathbf{p}})(\mathbf{p} - \hat{\mathbf{p}})^t \rangle, \quad (\text{A.2})$$

$$\mathbf{C}_s = \langle (\mathbf{s} - \hat{\mathbf{s}})(\mathbf{s} - \hat{\mathbf{s}})^t \rangle, \quad (\text{A.3})$$

where $\hat{\mathbf{p}}$ ($\hat{\mathbf{s}}$) denotes the estimated and \mathbf{p} (\mathbf{s}) the true values of primary (secondary) measure. Using (A.1) we can relate both covariances through

$$\mathbf{C}_s = \mathbf{A} \mathbf{C}_p \mathbf{A}^t. \quad (\text{A.4})$$

The transformation matrix \mathbf{A} has to be of rank $\mathbf{A} = n$, otherwise the covariance matrix of the secondary data vector $\mathbf{C}_s =$

$(\mathbf{A} \mathbf{C}_p \mathbf{A}^t)$ is singular and not invertible. Furthermore as \mathbf{A} is of dimension $(n \times m)$, rank $\mathbf{A} \leq m$ implying $n \leq m$. We take the χ^2 -functions a measure for the information content

$$\chi_p^2 = \Delta_p^t \mathbf{C}_p^{-1} \Delta_p \quad \text{and} \quad \chi_s^2 = \Delta_s^t \mathbf{C}_s^{-1} \Delta_s, \quad (\text{A.5})$$

where in our case $\Delta_p = \mathbf{p}^f - \mathbf{p}_\pi$ ($\Delta_s = \mathbf{s}^f - \mathbf{s}_\pi$) denotes the difference between the fiducial data vector \mathbf{p}^f (\mathbf{s}^f) and the data vector \mathbf{p}_π (\mathbf{s}_π) depending on the parameter vector π . In case χ^2 is minimal, the posterior likelihood of the corresponding π being the correct parameter vector is maximized. The difference between χ_p^2 and χ_s^2 characterizes which probability function has a larger curvature, i.e. which data vector gives tighter constraints in parameter space. Therefore the information content of primary and secondary data vector can be compared by calculating

$$\chi_p^2 - \chi_s^2 = \Delta_p^t \mathbf{C}_p^{-1} \Delta_p - \Delta_p^t \mathbf{A}^t (\mathbf{A} \mathbf{C}_p \mathbf{A}^t)^{-1} \mathbf{A} \Delta_p, \quad (\text{A.6})$$

for arbitrary Δ_p . In case this difference is always positive we can conclude that the primary data vector gives tighter constraints on parameters. We can always find transformation matrices \mathbf{V} (dimension $m \times m$) and \mathbf{U} (dimension $n \times n$) to rewrite the transfer matrix \mathbf{A} as an $n \times m$ matrix

$$\left(\mathbf{E}_n \middle| \mathbf{0} \right) = \mathbf{S} = \mathbf{U} \mathbf{A} \mathbf{V}^{-1} \quad \longleftrightarrow \quad \mathbf{A} = \mathbf{U}^{-1} \mathbf{S} \mathbf{V}. \quad (\text{A.7})$$

We can directly calculate these transformation matrices as a multiplication of elementary matrices (Fischer 1997a). Inserting (A.7) into (A.6) we derive after some lengthy but straightforward calculation

$$\chi_p^2 - \chi_s^2 = \Delta_p'^t \mathbf{C}'^{-1} \Delta_p' - \Delta_p'^t \mathbf{S}^t (\mathbf{S} \mathbf{C}' \mathbf{S}^t)^{-1} \mathbf{S} \Delta_p' \quad (\text{A.8})$$

with

$$\mathbf{C}' = \mathbf{V} \mathbf{C}_p \mathbf{V}^t \quad \text{and} \quad \Delta_p' = \mathbf{V} \Delta_p. \quad (\text{A.9})$$

For simpler notation we discard all “'” further on. We define

$$\mathbf{C}^{-1} = \left(\begin{array}{c|c} \mathbf{C}_1 & \mathbf{C}_2 \\ \hline \mathbf{C}_2^t & \mathbf{C}_3 \end{array} \right)^{-1} = \left(\begin{array}{c|c} \mathbf{D}_1 & \mathbf{D}_2 \\ \hline \mathbf{D}_2^t & \mathbf{D}_3 \end{array} \right), \quad (\text{A.10})$$

with \mathbf{C}_1 being an $n \times n$ matrix and calculate

$$\mathbf{S}^t \left(\mathbf{S} \mathbf{C} \mathbf{S}^t \right)^{-1} \mathbf{S} = \begin{pmatrix} \mathbf{C}_1^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}. \quad (\text{A.11})$$

Using (A.10) and (A.11) we can rewrite (A.8) as

$$\chi_p^2 - \chi_s^2 = \Delta_p^t \begin{pmatrix} \mathbf{D}_1 - \mathbf{C}_1^{-1} & \mathbf{D}_2 \\ \mathbf{D}_2^t & \mathbf{D}_3 \end{pmatrix} \Delta_p. \quad (\text{A.12})$$

From $\mathbf{C} \mathbf{D} = \mathbf{E}_m$ we deduce

$$\mathbf{C}_1 \mathbf{D}_1 + \mathbf{C}_2 \mathbf{D}_2^t = \mathbf{E}_n \longrightarrow \mathbf{D}_1 - \mathbf{C}_1^{-1} = -\mathbf{C}_1^{-1} \mathbf{C}_2 \mathbf{D}_2^t \quad (\text{A.13})$$

and

$$\mathbf{C}_1 \mathbf{D}_2 + \mathbf{C}_2 \mathbf{D}_3 = 0 \longrightarrow \mathbf{C}_2 = -\mathbf{C}_1 \mathbf{D}_2 \mathbf{D}_3^{-1}. \quad (\text{A.14})$$

Inserting (A.14) into (A.13) we can rewrite (A.12) as

$$\chi_p^2 - \chi_s^2 = \Delta_p^t \begin{pmatrix} \mathbf{D}_2 \mathbf{D}_3^{-1} \mathbf{D}_2^t & \mathbf{D}_2 \\ \mathbf{D}_2^t & \mathbf{D}_3 \end{pmatrix} \Delta_p. \quad (\text{A.15})$$

\mathbf{C} is positive definite and symmetric, therefore \mathbf{D}_3 as a submatrix is positive definite and symmetric and also the inverse \mathbf{D}_3^{-1} has these favorable properties (Anderson 2003). Hence, we can decompose $\mathbf{D}_3 = \mathbf{L} \mathbf{L}^t$ and finish our calculation as follows

$$\begin{aligned} \chi_p^2 - \chi_s^2 &= \Delta_p^t \begin{pmatrix} \mathbf{D}_2 (\mathbf{L}^t)^{-1} \\ \mathbf{L} \end{pmatrix} \underbrace{(\mathbf{L}^{-1} \mathbf{D}_2^t \mathbf{L})}_T \Delta_p \\ &= \Delta_p^t \mathbf{T}^t \mathbf{T} \Delta_p \\ &= \|\mathbf{T} \Delta_p\|^2 \\ &\geq 0. \end{aligned} \quad (\text{A.16})$$

We will now examine the case when $\chi_p^2 - \chi_s^2 = 0$. The information content of primary and secondary measure is considered to be equal if and only if this equality holds for **all** data vectors Δ_p . If there is only one Δ_p for which $\chi_p^2 - \chi_s^2 > 0$, the primary measure contains more information. The difference of the two χ^2 -values is given by (A.6). In case it is zero for all Δ_p ,

$$\mathbf{C}_p^{-1} = \mathbf{A}^t \left(\mathbf{A} \mathbf{C}_p \mathbf{A}^t \right)^{-1} \mathbf{A} \quad (\text{A.18})$$

must hold (Fischer 1997b). \mathbf{C}_p is of rank m , hence the lefthand-side of (A.18) must also have rank m . Then \mathbf{A} must have rank m and is therefore a quadratic $m \times m$ matrix, which is of course invertible. This result is intuitively clear, if one is able to calculate Δ_s from Δ_p and vice versa the information content should be the same. We can summarize the results of the above calculation in two statements:

1. If a secondary measure can be calculated from a primary by a matrix \mathbf{A} as described in (A.1), the secondary measure has less or equal information.
2. The amount of information is equal in case the rank of \mathbf{A} equals the dimension of the primary data vector (m) implying that \mathbf{A} is invertible.

References

- Anderson, T. W. 2003, An Introduction to Multivariate Statistical Analysis (Wiley-Interscience), 623/624
- Bacon, D., Refregier, A., & Ellis, R. 2000, MNRAS, 318, 625
- Bardeen, J. M., Bond, J. R., Kaiser, N., & Szalay, A. S. 1986, ApJ, 304, 15
- Bartelmann, M. & Schneider, P. 2001, Phys. Rep., 340, 291
- Crittenden, R. G., Natarajan, P., Pen, U.-L., & Theuns, T. 2002, ApJ, 568, 20
- Fischer, G. 1997a, Lineare Algebra (Vieweg), 161
- Fischer, G. 1997b, Lineare Algebra (Vieweg), 276
- Hartlap, J., Simon, P., & Schneider, P. 2007, A&A, 464, 399
- Hetterscheidt, M., Simon, P., Schirmer, M., et al. 2007, A&A, 468, 859
- Hoekstra, H., Mellier, Y., van Waerbeke, L., et al. 2006, ApJ, 647, 116
- Hoekstra, H., Yee, H. K. C., & Gladders, M. D. 2002, ApJ, 577, 595
- Jain, B., Seljak, U., & White, S. 2000, ApJ, 530, 547
- Jenkins, A., Frenk, C. S., White, S. D. M., et al. 2001, MNRAS, 321, 372
- Joachimi, B., Schneider, P., & Eifler, T. 2007, arXiv:0708.0387v1 [astro-ph]
- Kaiser, N. 1992, ApJ, 388, 272
- Kaiser, N. 1998, ApJ, 498, 26
- Kaiser, N. & Squires, G. 1993, ApJ, 404, 441
- Kaiser, N., Wilson, G., & Luppino, G. A. 2000, ArXiv Astrophysics e-prints
- Kilbinger, M. & Schneider, P. 2004, A&A, 413, 465
- Kilbinger, M., Schneider, P., & Eifler, T. 2006, A&A, 457, 15
- King, L. J. & Schneider, P. 2003, A&A, 398, 23
- Massey, R., Rhodes, J., Leauthaud, A., et al. 2007, ArXiv Astrophysics e-prints
- Ménard, B., Hamana, T., Bartelmann, M., & Yoshida, N. 2003, A&A, 403, 817
- Miller, K. 1981, Mathematics Magazine., 54, 67
- Schneider, P. 2006, Weak Gravitational Lensing (Springer Verlag, Berlin)
- Schneider, P., Kilbinger, M., & Lombardi, M. 2005, A&A, 431, 9
- Schneider, P., van Waerbeke, L., Jain, B., & Kruse, G. 1998, MNRAS, 296, 873
- Schneider, P., van Waerbeke, L., Kilbinger, M., & Mellier, Y. 2002a, A&A, 396, 1
- Schneider, P., van Waerbeke, L., & Mellier, Y. 2002b, A&A, 389, 729
- Schrabback, T., Erben, T., Simon, P., et al. 2007, A&A, 468, 823
- Semboloni, E., Mellier, Y., van Waerbeke, L., et al. 2006, A&A, 452, 51
- Smith, R. E., Peacock, J. A., Jenkins, A., et al. 2003, MNRAS, 341, 1311
- van Waerbeke, L., Mellier, Y., Erben, T., et al. 2000, A&A, 358, 30
- van Waerbeke, L., Mellier, Y., & Hoekstra, H. 2005, A&A, 429, 75

Wittman, D. M., Tyson, J. A., Kirkman, D., Dell'Antonio, I., & Bernstein, G. 2000, *Nature*, 405, 143

Acknowledgements. We thank Yannick Mellier, Jan Hartlap and Tim Schrabback for useful discussions and advise. This work was supported by the Deutsche Forschungsgemeinschaft under the projects SCHN 342/6-1 and SCHN 342/9-1. TE is supported by the International Max-Planck Research School at the University Bonn. MK is supported by the CNRS ANR “ECOSSTAT”, contract number ANR-05-BLAN-0283-04.